XI. Statistical Properties of Large Scale Structure

The characterization of the distribution of galaxies and mass in the universe is a well developed field, and only a few basic priciples will be outlined here. More extensive discussions can be found in books by Peebles and others.

A. Two Point Correlation Function. RMS fluctuations in galaxy counts, mass in spheres

There are multiple ways to characterize the statistical properties of mass fluctuations, all of which (at the level of detail that we are interested in) are equivalent to one another. All methods capture the fact that mass is not distributed randomly but rather that fluctuations are correlated, and the amplitude of the fluctuations is larger on smaller scales.

We will describe three characterizations here: 1) the correlation fuction $\xi(r)$, b) the rms fluctuations $(\Delta M/M)_{rms}$ on mass scale M, and c) the power spectrum $P(k)$.

1. Correlation function

Let $\delta(r) = \Delta\rho/\bar{\rho} - 1$ be the (dimensionless) density fluctuation field relative to the mean density$\bar{\rho}$. We note that δ is a function of time, but will suppress that dependence for now.

We define the correlation function $\xi(r)$ to be

$$\xi(r) = \langle \delta(r')\delta(r'+r) \rangle, \tag{11.1}$$

or

$$\xi(r) = \frac{1}{V}\int \delta(\vec{r'})\delta(\vec{r'}+\vec{r})dV', \tag{11.2}$$

where the expected value represents an average over r'. Formally, r and r' are vectors, and the average is one over 3 dimensions. For an isotropic universe, $\xi(r)$ depends only on the magnitude of r, not direction, and thus is a 1 dimensional function. On scales of a few Mpc the universe today has a correlation function that is nearly a power law:

$$\xi(r) \approx (r/r_0)^{\gamma}. \tag{11.3}$$

Observations show that power law index $\gamma \approx -1.8$ and the correlation length $r_0 \approx 5h^{-1}$ Mpc, although the length depends on galaxy type.

2. Mass fluction spectrum.

For any volume of space encompassing an average mass M, the fractional excess (or deficit) of mass is given by

$$\frac{\Delta M}{M}(\vec{r}) = \left(\frac{1}{V}\right)\int_0^a \delta(\vec{r}+\vec{r'})dV', \tag{11.4}$$

where $\vec{r}$ is the location in space of the volume and a is the radius of the volume. We are interested in the variance of this quantity:

$$\left\langle \left(\frac{\Delta M}{M}\right)^2 \right\rangle. \tag{11.5}$$

We have

$$\left(\frac{\Delta M}{M}\right)^2 = \left(\frac{1}{V'}\right)\left(\frac{1}{V''''}\right)\int_0^a \int_0^a \delta(\vec{r'} + \vec{r})\delta(\vec{r} + \vec{r'})dV'dV''.. \tag{11.6}$$

We average this quantity over all of space:

$$\begin{aligned} \left\langle \left(\frac{\Delta M}{M}\right)^2 \right\rangle =& \left(\frac{1}{V}\right)\left(\frac{1}{V'}\right)\left(\frac{1}{V''''}\right)\int_0^a \int_0^a \delta(\vec{r'} + \vec{r})\delta(\vec{r} + \vec{r'})dV'dV''dV \\ =& \left(\frac{\Delta M}{M}\right)^2 = \left(\frac{1}{V'}\right)\left(\frac{1}{V''''}\right)\int_0^a \int_0^a \xi(\vec{r'} - \vec{r})dV'dV''''. \end{aligned} \tag{11.7}$$

The last double integral is weighted average of the correlation function $\xi(a)$ inside a sphere of radius a; for our purposes, it is sufficient to know that this is approximately equal to the correlation function evaluated at radius a.

3. Power spectrum.

It is simplest to peform a 1-dimensional calculation; the generalization to 3 dimensions is straightforward.

Consider a portion of space of length L. Let x be the space coordinate such that $-L/2 < x < L/2$. Let $\delta(x)$ be the dimensionless perturbed density. We can express $\delta(x)$ as a Fourier series:

$$\delta(x) = \sum_{n=0}^{\infty} A_n \cos(\frac{2\pi n}{L}) + B_n \sin(\frac{2\pi n}{L}). \tag{11.8}$$

If we introduce $\alpha = \frac{1}{2}(A_n + iB_n)$ and extend the sum to $-\infty$, we can write

$$\delta(x) = \sum_{n=-\infty}^{\infty} \alpha_n \exp(-2\pi i x n/L). \tag{11.9}$$

The coefficients are given by:

$$\alpha_n = \frac{1}{L}\int_{-L/2}^{L/2} \delta(x)\exp(2\pi i n x/L)dx. \tag{11.10}$$

We seek to convert this to a integral formulation. Let $k = n/L$ and $F(k) = \sqrt{L}\alpha(Lk)$. (The normalization of F is chosen to simplify the power spectrum calculation later). The summation can be converted to an integral:

$$
\begin{aligned}
\delta(x) =& \sqrt{L}\int_{-\infty}^{\infty} F(k)\exp(-2\pi ikx)dk && (11.11)\\
F(k) =& \frac{1}{\sqrt{L}}\int_{-L/2}^{L/2} \delta(x)\exp(2\pi ikx)dx. && (11.12)
\end{aligned}
$$

The correlation function can be expressed as

$$
\begin{aligned}
\xi(z) =& \frac{1}{L}\int_{-L/2}^{L/2} \delta(x)\delta(x+a)dx\\
=& \int_{-L/2}^{L/2} dx \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(k)F(k')\exp(-2\pi ikx)\exp(-2\pi ik'[x+a])dkdk' && (11.13)
\end{aligned}
$$

The integral from $-L/2$ to $L/2$ can now be extended to $\pm\infty$ so the integral over x becomes a delta function of $k+k'$. Setting $k' = -k$, we arrive at

$$
\xi(a) = \int_{-\infty}^{\infty} |F(k)|^2 \exp(2\pi ika)dk. \qquad (11.14)
$$

In three dimensions we get

$$
\xi(\vec{r}) = \int |F(k)|^2 \exp(2\pi i\vec{k}dot\vec{r})d^3k. \qquad (11.15)
$$

Only the real part of the complex exponential survives. If we align k_z with $\vec{r}$ and convert the k space to polar coordinates, we have $\vec{k}dot\vec{r} = kr\cos(\theta)$ and $d^3k =>$ $k^2\sin(\theta)d\theta d\phi$. Integrating over θ and ϕ, we have

$$
\int \cos(2\pi kr\cos\theta)\sin(\theta)d\theta d\phi = \frac{\sin(2\pi kr)}{kr}. \qquad (11.16)
$$

Thus,

$$
\xi(r) = \int_0^{\infty} |F(k)|^2 \frac{\sin(2\pi kr)}{kr}k^2 dk. \qquad (11.17)
$$

This completes the computation of the inter-relationships among the three formulations of density fluctuations.

B. Initial Fluctuation Spectrum

A useful approximation to the power spectrum is to assume that it has the form of a power law: $P(k) \propto k^n$. Insert this form into Eq. (11.17) and integrating, we find

$$
\xi(r) \propto \frac{1}{r}k_{max}^{n+2}. \qquad (11.18)
$$

The integral is oscillatory for large k. There are standard "tricks" for taming such behavior, and the result is that the integral is finite with $k_{max} \approx 1/r$. Thus,

$$\xi r \approx r^{-(3+n)}. \tag{11.19}$$

The mass fluctuation have the form

$$\left(\frac{\Delta M}{M}\right)^2 \propto r^{-(3+n)} \propto M^{-(3+n)/3} \tag{11.20}$$

(since $M \propto r^3$). The rms fluctuation is thus

$$\sigma_M \propto M^{-(n+3)/6}. \tag{11.21}$$

The Harrison-Zel'dovich hypothesis holds that the rms mass fluctuation for massses on the size of the horizon at any particular time is a constant κ. Once inside the horizon the perturbations grow in size due to gravitational instability. We would like to calculate the shape of the power spectrum today. The calculation of the growth factor in general is quite complex for perturbations that turn into things that we see today such as galaxies and clusters because perturbations of those scales entered the horizon during the radiation dominated epoch, when baryons are tightly coupled to the radiation and effects such as radiation damping are important. However, it is straightforward to calculate the growth factor for large scale perturbations that entered the horizon during the matter dominated era.

The calculation proceeds as follows. Perturbations larger than the horizon size do not grow at all. Once a perturbation enters the horizon it starts growing like $\delta \propto t^{2/3}$, as we showed in the last chapter. Today, at time t_0, the rms amplitude of a perturbation on mass scale M is given by $\sigma_M = \kappa(t_0/t_H)^{2/3}$, where t_H is the time that that particular mass scale entered the horizon.

At any time the radius to the horizon is proptional to H^{-1}, The density $\rho \propto H^2$, and the mass $M_H \propto \rho r^3 \propto H^{-1}$. For a critically bound universe, $H \propto t_H^{-1}$, so $M_H \propto t_H$. Thus, today, the mass spectrum is $\sigma_M \propto M^{-2/3} = M^{-(n+3)/6.}$ Solving for n, we find $n = 1$. This spectrum is referred to as the "initial power spectrum" and predicted by many standard models of inflation.

For lower mass perturbations that arose during the radiation era, growth is suppressed due to the dominance of the radiation, which propagates perturbations as sound waves rather than allowing large growth. The physics will not be pursued here, but they are described in books such as Kolb and Turner, and in Dodelson (??).

C. Press-Schechter Theory

While the statistical description of mass fluctuations in the early universe is relatively straightforward, the connection between small linear fluctuations at that time and the highly nonlinear perturbations in the current universe is not trivial to compute. Press and Schechter (1974) developed a simple methodology to characterize the number distribution of bound objects as a function of mass that qualitatively seems to describe the luminosity function of galaxies and mass function of galaxy clusters reasonably well.

Consider a volume of universe V with mass density today of ρ_0. In current cosmological models that involve dark energy, the mass density refers only to that of dark matter and other mass components that are cold and and cluster in bound systems. Consider the state of such a volume at an early time t_i when the density fluctuations were small. Divide the volume into a number of equal sized regions with mean mass per region of M and true total mass in region i of $M(1+\delta_i)$. Let $F(M,\delta)$ be the fraction of regions measured on mass scale M with mean overdensity $\leq \delta$. In a universe with a Gaussian distribution of fluctuations, the function F is given by

$$\frac{\partial F}{\partial \delta} = \frac{1}{\sqrt{2\pi}\sigma_M}\, e^{-\frac{\delta_M^2}{2\sigma_M^2}} \tag{11.22}$$

where σ_M is the rms value of δ on mass scale M. One can repeat the exercise for different mass scales M. The normalization is chosen such that $F = 1$ when $\delta = \infty$. As the volume of universe evolves, overdense regions grow, and those regions with overdensities greater than some δ_{crit} collapse. On a given mass scale M, the fraction of regions that collapse is given by

$$F_{collapsed} = \frac{1}{\sqrt{2\pi}\sigma_M}\int_{\delta_{crit}}^{\infty} e^{-\frac{\delta_M^2}{2\sigma_M^2}}\, d\delta \tag{11.23}$$

$$= \frac{1}{\sqrt{2\pi}}\int_{\frac{\delta_{crit}}{\sigma_M}}^{\infty} e^{-\frac{x^2}{2}}\, dx. \tag{11.24}$$

One can repeat the same calculation on a slightly larger mass scale $M+\Delta M$. For any sensible mass spectrum, the fraction of regions that have collapsed on the larger scale is slightly less than that on the smaller scale. Note that these calculations are simply different characterizations of the same physical volume. Most of the collapsed regions on mass scale M are not intact but are subsumed into slight larger regions of mass $M+\Delta M$. Press and Schechter argue that the small fraction of regions of mass M not subsumed into larger regions should survive as intact objects. The number of such regions is given by

$$\Delta n = -[F(\delta_{crit}, M+\Delta M) - F(\delta_{crit}, M)]\frac{\rho_0}{M}, \tag{11.25}$$

where n has units of number of objects per volume. With a bit of manipulation of the equation for F, this equation becomes

$$\frac{dn}{dM} = -\frac{dF}{dM}\frac{\rho_0}{M}$$

$$= -\frac{\rho_0}{M}\frac{1}{\sqrt{2\pi}}\frac{\delta_{crit}}{\sigma_M^2}\frac{d\sigma_M}{dM}\; e^{-\frac{\delta_{crit}^2}{2\sigma_M^2}}. \tag{11.26}$$

The above process applies only to half the mass in the volume, that in overdense regions. The underdense regions will empty out and pile up mass on surrounding shells. Press and Schechter argue (not entirely convincingly) that this mass will also form into bound regions and have a number distribution like the overdense regions above, such that the total number density is actually twice that in the above equation. Thus, the final number density is given by

$$\frac{dn}{d\ln M} = -\frac{\rho_0}{M}\sqrt{\frac{2}{\pi}}\frac{\delta_{crit}}{\sigma_M}\frac{d\ln\sigma_M}{d\ln M}\; e^{-\frac{\delta_{crit}^2}{2\sigma_M^2}}. \tag{11.27}$$

To proceed further we need an expression for the amplitude of the perturbation spectrum σ_M and we need to tie the critical overdensity δ_{crit} to some observable quantity today. Inflation theories predict that the initial fluctuation spectrum has a nearly power law dependence of the form:

$$\sigma_M \propto M^{-(3+n)/6} = M^{-\alpha}, \tag{11.28}$$

where n is a constant characterizing the initial power spectrum of density fluctuations. Again, inflation theories predict $n \approx -1$.

To tie the amplitude to an observable quantity today, it is first necessary to return to the concept of linear growth factor. Small amplitude density perturbations grow in time at a rate that is the same regardless of the initial amplitude or mass scale of the perturbation. Equation 10.15 gives an explicit expression for this growth rate in a critically bound matter-dominated universe. In general it is conventional to define a function $D(z)$ that gives the ratio of the size of a linear perturbation at the time corresponding to redshift z to the size of that perturbation today. For a critically bound matter dominated universe, we have $D(z) = (1+z)^{-1}$.

The standard convention for defining the amplitude of the perturbation spectrum is to use a quantity called σ_8, which is almost (but not quite) the rms density fluctuation today in spheres of radius $8h^{-1}$ Mpc. Observationally the true density fluctuation on this scale is about unity, which means that one cannot quite use the linear growth factor $D(z)$ to express the amplitude of the density fluctuations at any time in the past. Let M_8 be the mass inside the $8h^{-1}$ Mpc sphere and $\sigma_8(z)$ be the rms amplitude of density fluctuations on this mass scale at an early epoch of the universe when such fluctuations are small enough that the linear approximation holds. σ_8 is defined to be the amplitude that these perturbations would have today assuming that the linear approximation were still valid: $\sigma_8 = \sigma_8(z)/D(z)$. The difference between this extrapolated value and the true rms value of density perturbations on this mass scale today is small (less than 10%) and depends somewhat on the exact cosmology.

Return now to equation xxx for the Press-Schechter spectrum. It is convenient define a mass scale M_* for which a positive density perturbation of exactly 1 σ is just collapsing today. By Eq. (10.15), we see that in the linear approximation, such a region would have an overdensity of 1.69 today. We can thus express M_* in units of σ_8 and M_8 as follows:

$$\frac{1.69}{\sigma_8} = \left(\frac{M_*}{M_8}\right)^{-\alpha}, \tag{11.29}$$

or

$$M_* = M_8 \left(\frac{\sigma_8}{1.69}\right)^{1/\alpha}. \tag{11.30}$$

With the above definition of M_*, we immediately find that the critical overdensity δ_{crit} in Eq. (XX) is given by $\delta_{crit} = \sigma(M_*)$. At some earlier epoch, the required overdensity is larger, and we have $\delta_{crit}(z) = \sigma(M_*)/D(z)$ (with M_* still being its present day value). With this generalization, the Press Schechter mass function can finally be written in the form:

$$\frac{dn}{d\ln M} = \frac{\rho_0}{M_*}\sqrt{\frac{2}{\pi}}\frac{M_*}{M}\frac{1}{D(z)}\left(\frac{M}{M_*}\right)^{\alpha} \alpha \; e^{-\frac{1}{2D(z)^2}\left(\frac{M}{M_*}\right)^{2\alpha}}. \tag{11.31}$$

In this equation, ρ_0 and M_* refer to values at $z = 0$.